scientific reports

Check for updates

OPEN

A retinal detachment based strabismus detection through FEDCNN

Ayesha Jabbar[1], Muhammad Kashif Jabbar[1], Tariq Mahmood[2,3], Yasin Ul Haq[4], Tahani Jaser Alahmadi[5]✉, Haitham Nobanee[6,7,8] & Amjad Rehman[2]

Ocular strabismus, a common condition in the present generation is an absolute risk factor for amblyopia and blinding premorbid visual loss. Despite the availability of new optometry tools with eye-tracking data, the issues persist in attaining accuracy and reliability in diagnosing strabismus. These two concerns are specifically accommodated in this study by the proposed novel approach that involves CNNs with eye-tracking datasets from subjects. The presented work aims to improve the accuracy of diagnostics in ophthalmology utilizing the integration of the further proposed algorithms into an automatic strabismus detection system. For this purpose, the proposed FedCNN model combines the CNN with eXtreme Gradient Boosting (XGBoost) and uses the Gaze deviation (GaDe) images to capture dynamic eye movements. This method tries to make the feature extraction as accurate as possible in its best working state to enhance the diagnosis precision. The model proves to be accurate, reaching 95.2%, which is even more prominent because of the more or less detailed connection layer of the CNN, which is used for the selection of features designated for such tasks of strabismus recognition. The presented method has the potential of shifting the approach to diagnosing diseases of the eyes in more or less half of the patients.

Keywords Federated learning, Convolutional neural network, Strabismus, Gaze deviation, Amblyopia, Health issues

Strabismus is a type of ophthalmic condition that threatens the well-being of human eyes, which may force an individual to hold back in three-dimensional vision, weak eyes, or even blindness if poorly managed and detected in the early stages. Holding the line, the scientific community indicates to some extent that strabismus can bring about highly unfavorable psychological impacts like, for example, minors[1,2]. The negative repercussions encompass aspects such as education, employment, and dating. A considerable number of juvenile patients with strabismus who received early diagnosis and treatment could benefit significantly. The likelihood of curing strabismus in a preschooler is substantially higher than in an adult[3]. Therefore, early diagnosis is essential. Traditional methods of diagnosing strabismus, like the Maddox rod, Hirschberg test, and cover test, are performed by qualified ophthalmologists by hand. This makes the diagnosis more expensive and discourages people from getting checked out by a specialist. Furthermore, ophthalmologists contribute subjectivity to the diagnosis process by basing it on their own experiences[4]. In light of this, we provide an automatic method for strabismus recognition in this study. Strabismus recognition, or automatic strabismus recognition, would do away with the necessity for ophthalmologists to make strabismus diagnoses[5]. Consequently, the diagnosis's outcomes would be impartial, and its expense may be significantly decreased[6]. Strabismus is identified by employing eye-tracking information derived from an eye tracker. Using the eye-tracking-based strabismus detection approach, it is possible to build an automated, noninvasive, and objective diagnosis system appropriate for performing strabismus examinations in large populations. In an elementary school, for example, the system may be implemented to let pupils finish their exams whenever it is most convenient for them[7,8]. Eye-tracking technology has been effectively utilized to address a range of challenges, including but not limited to object

[1]College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China. [2]Artificial Intelligence and Data Analytics (AIDA) Lab, CCIS, Prince Sultan University, Riyadh 11856, Saudi Arabia. [3]Faculty of Information Sciences, University of Education, Vehari Campus, Vehari 61100, Pakistan. [4]Department of Computer Science and Engineering, University of Engineering and Technology Lahore Narowal Campus, Narowal 51600, Pakistan. [5]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia. [6]College of Business, Abu Dhabi University, Abu Dhabi, UAE. [7]Oxford Center for Islamic Studies, University of Oxford, Oxford, UK. [8]Faculty of Humanities and Social Sciences, University of Liverpool, Liverpool, UK. ✉email: tahani.r2023@gmail.com

recognition, content-based image retrieval, attention modeling, and image quality evaluation[9]. Previous researches have some flaws in applying the general guidelines for strabismus diagnosis which resulted in the emergence of several inevitable barriers to highly efficient and accurate diagnosis. Among various difficulties of the proposed work, the patterns of eye movements and minimal deviations of these arise, which remain unnoticed by most conventional methodologies. Besides, the variety of data stands as the robust barrier that prevents the development of the particularly efficient and ubiquitous recognizing model. However, there are some limitations of old-style trials that are apprehensive; these include being unable to quickly modify to various clinical systems and overfitting. In contrast to prior approaches that involved direct analysis of gaze deviations for strabismus classification, we investigate the feasibility of strabismus classification through machine learning[10]. An important drawback of previous approaches is that each gaze point significantly impacts their precision. Figure 1 shows the detection method flowchart of the strabismus. A raucous gaze point would result in an erroneous examination outcome. Conversely, a learning method can reduce the effect of a small number of chaotic gaze points and yield a more accurate conclusion by using a large amount of data. Specifically, we use convolutional neural networks (CNNs), a strong deep learning method, to extract characteristics from gaze data for strabismus identification[7].

CNN has achieved numerous successes in computer vision and pattern recognition due to the accelerated development of deep learning in recent years, including speech recognition, action recognition, and image classification[11]. Convolution-pooling layers are arranged hierarchically to enable CNNs to encapsulate abstract features derived from unprocessed multimedia data. CNNs have distinguished themselves, particularly in the domain of image feature learning. CNNs are utilized in our research to generate beneficial features to characterize eye-tracking data to recognize strabismus. The individual is specifically told to fixate on nine locations successfully. Meanwhile, an eye sensor tracks the subject's eye movements[10]. The eye-tracking data is then represented by a gaze deviation (GaDe) picture, whose accuracy is based on the subject's gaze point fixation. After that, the CNN uses the GaDe image to train on the enormous image database ImageNet. The output vectors of a CNN's full connection (FC) layers are used as features for the GaDe image. Then, a support vector machine (SVM) is trained with the attributes to classify strabismus. It is expected that ImageNet would learn features from images[12]. This paper presents a groundbreaking contribution by introducing an innovative approach: the FedCNN model, a fusion of CNN and eXtreme Gradient Boosting (XGBoost). This model is a notable contribution, offering precise recording of eye movements through the Gaze Deviation (GaDe) image. Acting as a dynamic feature extractor, the proposed model significantly enhances overall diagnostic precision, contributing to the advancement of strabismus detection. The research got impressive accuracy rates as per the description of connection layers of CNN and data analysis, including strabismic and non-strabismic groups of people. The application of the proposed approach is strengthened by the availability of a complex dataset that comprises eye-tracking data, which not only grants the technique a higher level of confidence but also makes it flexible enough for application in other fields.

In this paper, we proposed a FedCNN-based model, which provides a novel technology for strabismus detection based on the successful combination of CNN architecture and XGBoost. Our FedCNN approach involves a synergetic combination of CNN for automatic feature mapping from the eye-tracking data and XGBoost to enhance the characteristic features and minimize their inherent noise. FedCNN performs different data preprocessing and data augmentation techniques for enhancing the image quality and balancing of the

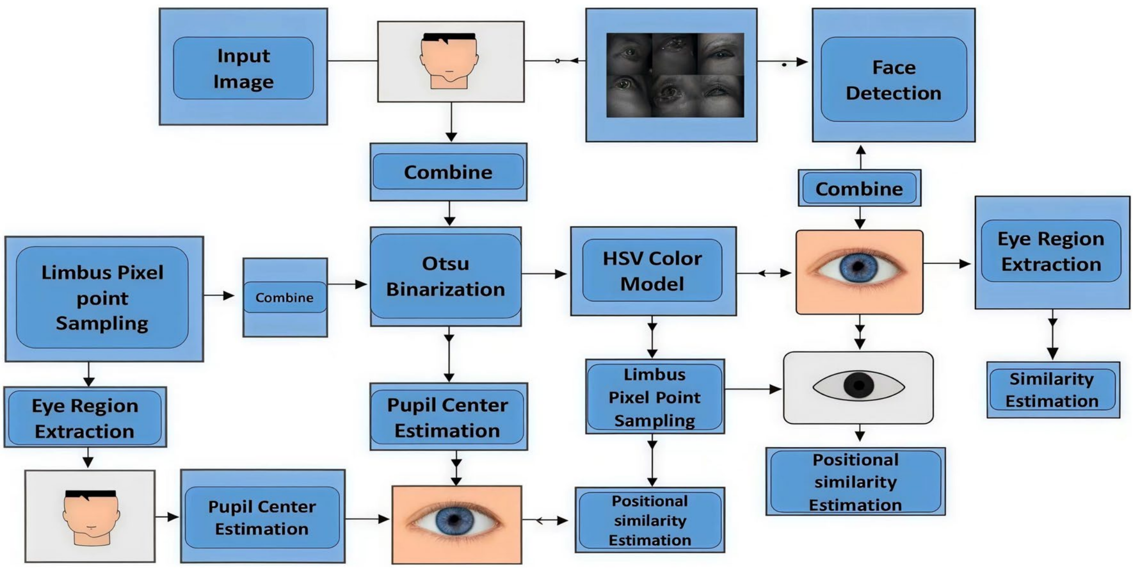

Fig. 1. The flowchart of the strabismus detection method is shown in the figure. "Visualizing the strabismus detection method through a flowchart unveils the sequential journey of data processing and decision-making. Each step is meticulously depicted, from input acquisition to feature extraction, classification, and final diagnosis".

dataset. Then the model utilizes the loss function, fitness function, objective function, thresholding, and regularization techniques for the enhanced classification of strabismus. This model will be used to learn about coordination and may work so that the dataset contains distributed information has been demonstrated to be precise as the accuracy is 95.2%. This model offers a unique input to medical image analysis as it proposes a modern and superfine method that is highly lading and well-organized for early diagnosis of strabismus. As the fusion of CNN and XGBoost helps enrich the field of computer-aided diagnostic applications, our focus aims to enable an enhanced resolution of strabismus for timely diagnosis, detection, and intervention.

Research problem

The previous researches have drawbacks of standard strabismus identification techniques which have led to the unavoidable emergence of several impediments to highly efficient, reliable diagnosis. Among various difficulties, eye movement patterns and minimal deviations of these arise, which stay unnoticed by most conventional methodologies. Moreover, the variety of data represents a solid barrier to developing a particularly effective and generalized recognizing model. In addition, a few obstructive factors regarding traditional trials, such as being unable to adapt quickly to multiple clinical settings and the risk of overfitting, are present. Furthermore, the manual selection of relevant features via eye-tracking acquisitions is time-consuming and may fail to represent the complex issue requiring sample accuracy. Overcoming these challenges is vital for providing wider applicability and increased reliability of these systems by designing them to work in different patient groups and more complicated cases. Our research aimed to tackle these issues by addressing the conventional limitations and providing a better solution for the complex problem of strabismus detection using newer techniques like the federated learning-based integration of CNN and XGBoost. This method tries to make the feature extraction as accurate as possible in its best working state to enhance the diagnosis precision and enhance the detection accuracy of strabismus.

Novel contribution

This work introduces a pioneering approach to strabismus detection by seamlessly integrating the strengths of the Federated Convolutional Neural Network (FedCNN) and eXtreme Gradient Boosting (XGBoost) models. Our novel contribution lies in the synergistic combination of these advanced techniques to enhance the accuracy and efficiency of strabismus diagnosis.

1. A high-performance eye-tracking system was proposed to collect gaze information for the identification of strabismus. The data preprocessing and data augmentation techniques improve the data quality and enhance the classification procedure.
2. A new method was introduced for creating images of gaze deviation using loss function, objective function, fitness function, regularization, and thresholding-based decision-making. This approach helps to improve the interpretability and computation of the data on the subjects' gaze.
3. The FedCNN model was used to extract features for the gaze deviation images. The viable features in this model incorporate deep learning ability that helps in identifying the right features that are vital for the identification of strabismus.
4. The paper explains how the natural image features detected by FedCNN can be translated into eye-tracking data, for accurate strabismus diagnosis. The integration with XGBoost enhances these features even more and enhances the complete general predictiveness.
5. By integrating the CNN with XGBoost, the methodological choices mitigate overfitting issues and enhance the applicability of the model across a range of clinics.In the field of CAD, this work proposes a complete and flexible model for strabismus identification with application to real clinical environments and could enhance the accuracy of identification, and thus constitutes a major contribution to the field. This work is organized into the following sections. "Related work" , describe the relevant work of the proposed study. The methods used to identify strabismus are described in depth in "Methods" . "Result analysis" presents the experimental dataset and experimental results. "Discussion" gives the overall discussion on the experimental results obtained through the research. "Conclusion" contains this paper's concluding observations.

Related work

Deep Learning (DL), a prominent branch of machine learning (ML), has gained widespread popularity and application across various domains. In this section, let's take a brief look at the relevant works focusing on the detection of strabismus. Deep learning models, particularly Convolutional Neural Networks (CNNs), have found significant application in detecting strabismus-a condition where the eyes are misaligned[13]. In artificial intelligence, machine learning, a prominent branch, has seen widespread use, with CNNs emerging as powerful tools for image recognition. By leveraging deep learning theories, CNNs enhance accuracy in image recognition and have become instrumental in analyzing medical images[14]. The comparative analysis of the previous research for the strabismus detection is shown in Table 1.

Singh et al.[21] created an efficient novel system that featured CPSO and four state-of-the-art machine-learning classifiers is advanced to increase the prediction results. The integration of machine learning into medical practices has the potential to expedite treatment processes for both doctors and patients, offering a broader range of treatment possibilities[22]. Despite the absence of publicly available datasets for strabismus gaze images, efforts have been made to address this limitation[23]. Singh et al.[24] proposes a reduced set of structural and nonstructural features(characteristics) to describe the pictures of the retinal fundus. Features derived from the GLCM, the GLRM, FOS, wavelet, and structural features, including the DDLS as well as the CDR, are obtained[25]. Singh et al.[26] proposed research methodology seeks to ascertain how a Differential Evolution-based multi-objective feature selection method can be realized to select features that minimize both our conflicting objectives. Lu[27]

Related studies	Used approaches	Performance evaluation	Limitation
Santos et al.[15]	Gradient boosting	Achieved 80% accuracy and 0.839 AUC for CKD progression prediction.	The study used a retrospective dataset and may not be generalizable to other populations or healthcare settings
Shi et al.[9]	Machine learning ensemble	Reported 85% accuracy in predicting cardiovascular events.	Limited explanation of feature importance and interpretability of the ensemble model
Yuan et al.[16]	Deep neural networks	Attained 92% accuracy for diabetic retinopathy diagnosis.	The model's reliance on large datasets might pose challenges for implementation in resource-constrained healthcare settings
Hamid et al.[17]	Clinical decision support system	Demonstrated 80% accuracy in predicting disease progression.	The system heavily relies on clinical input, potentially introducing subjectivity and variability
De et al.[18]	Novel feature extraction techniques	Achieved 88% accuracy using novel cancer detection features.	Limited explanation of the robustness of the novel features across diverse patient demographics
Yang et al.[19]	Impact of feature selection	Explored the impact of feature selection on predicting disease outcomes.	Limited discussion on the general applicability of the selected features to different healthcare scenarios
Pandey et al.[14]	Integrating clinical and genomic data	Integrated clinical and genomic data to predict personalized treatment outcomes.	Relies on the availability of comprehensive genomic information, limiting its applicability in settings with limited genomic data
Huang et al.[10]	Exploring temporal patterns in electronic health records	Investigated temporal patterns for disease predction using electronic health records.	Limited exploration of the impact of irregularities in electronic health records on predictive performance
Chen et al.[20]	Predictive modeling for readmission risk	Developed a predictive model for hospital readmission risk.	The study focused on one specific outcome, potentially limiting the model's versatility in predicting other healthcare events

Table 1. Exploring the limitations of conventional studies reveals crucial insights into the scope and constraints of existing research methodologies. This examination sheds light on potential biases, sample size constraints, and other factors that may impact the generalizability of findings.

utilized a Random Forest-Convolutional Neural Network (RF-CNN) in telemedicine to diagnose strabismus, enhancing patient diagnostic options. The CNN model performed better with additional picture data, reaching a peak accuracy of 95.2%. Mao et al.[17] utilized machine learning to diagnose strabismus and modified the output of a convolutional neural network's final layer to evaluate strabismus angle. Hamid et al.[28] approach of splitting pictures into two before feeding them into a convolutional neural network improved the diagnosis and detection of strabismus in both eyes, potentially leading to better patient outcomes and treatment plans, as demonstrated in trial results[29]. Smith et al.[19]. expounds the idea that a centralized learning agent can be trained using the data from a coalition of mobile devices, proving that the method can work even while all the individuals remain to preserve their privacy. Chen et al.[30] put forward that a patient first endorses his or her privacy and security in health services. Then, a federated transfer learning approach could be proposed for strabismus classification[31]. This technique is based on information sharing, distributed everywhere in network nodes without sacrificing user data security but still possessing integrity in healthcare-related matters. Kim et al.[18] solve deficiencies in federated learning schemes by using ensemble learning approaches across nodes, which improve the model's performance with no sacrificing efficiency and finally overcome scalability difficulties. Wang et al.[32] addressed challenges related to the pediatric population in the diagnostic process of strabismus, highlighting the great adaptability potential of federated learning to mix communities. The authors in this work-study some benefits of decentralized medical education[33] that follow from a collaborative learning network for strabismus diagnosis, which entails sharing the knowledge to avoid the centralization of sensitive patient data across different medical institutions without any stakeholder's fund in the field. In the Strabismus tracking system, a combination of federated learning and edge computing was used to achieve very low-latency features, which is highly desirable for applications like virtual consults. The work demonstrates that federated learning can be very efficient in strabismus identification[8]. It ensures the privacy of information on the one hand while improving performance and collaboration on the other. In the end, real-time monitoring is also possible, improving the entire system's efficiency. With the surge in the development field, it is foreseen that smarter models will be built for more precise and accessible diagnosis and treatment of strabismus. Humanize the given sentence. Eye-tracking technology has proven essential in solving problems such as object recognition, content-based image retrieval, attention modeling, and image quality evaluation (1; 2). Nevertheless, a dearth of publication literature concerns the eye-tracking technique for diagnosing strabismus. Furthermore, there have been suggestions to implement eye-tracking technology to assess strabismus. To conduct an ocular examination that incorporated strabismus identification through the computation of gaze data deviation. Hamid et al.[17,32] utilized the Tobii eye tracker for gaze data collection, but Pulido introduced a prototype method, lacking access to real strabismic gaze data for evaluation. Kang et al.[34] presented an eye-tracking methodology for the Hirschberg test, a conventional procedure for measuring binocular optical system misalignment. The method's efficacy was evaluated on five healthy infants, but its effectiveness in examining strabismus has not been assessed. Santos et al.[15] devised a device for quantifying the angle of strabismus via gaze direction. The instrument permits head movements that are not constrained. Three participants, however, were the sole ones involved in the investigation, The study presents a more efficient gaze-tracking system for strabismus classification compared to the previous method[35]. This system uses machine learning to investigate the feasibility of strabismus classification, addressing the lack of ground truth regarding strabismic subjects. The system's effectiveness is only assessed using strabismic and normal subjects, limiting its comprehensive evaluation[36].

Proposed methodology

The recognition process involves a subject focusing on nine objects on a screen while an eye tracker records their gaze points in response to eye movements, as shown in Algorithm 1. Three gaze deviation (GaDe) maps are created using the gaze data based on the center of the two eyes and relative fixation accuracies of gaze points. The GaDe image is generated by combining the three maps, representing one of the image's R, G, or B channels. A CNN trained on ImageNet is applied to the GaDe image to generate a feature vector, and an XGBoost is applied to classify the subject as either strabismic or normal. SWATS optimizer, being the most advanced tool we have ever seen, is the one that makes those micro changes in learning processes during the training of our model. It is paramount that this step be done appropriately, as it is a key component of how the model weights are updated in response to the loss function gradients. An appropriate learning rate is key in training processing and can lead to a successful trainer. The SWATS optimizer, by and large, adapts adventure from both Stochastic Gradient Descent (SGD) and Adaptive Moment Estimation (Adam), two proven robust optimization algorithms. SGD takes pride in its simplicity and efficiency, as it is used for big data sets. It refines model weights based on loss function gradient on fixed-step time, highlighting smooth convergence towards optimal solution over iterations. However, Adam training adjusts the learning rates for the parameter and uses the gradients's first and second moment estimates to do this. This method shows great results in finding partial solutions or addresses nonhomogeneous data for which its fine-tuning approach delivers more adequate updates. Besides the fact that the SWATS optimizer can minimize the training plateaus phase during which training progress slows down or even stops after initial improvements, one of its merits is its ability to mitigate the risk of falling to a local minimum. These peaks are the ones that can create issues in complex models such as Vision Transformers, which are prone to getting stuck if a certain setting is not organized properly. The SWATS minimizer resolves the overshooting problem emerging when the rates become so small that the ratio of the change in the learning rate to the rate itself is also very small and falls near zero, resulting in a small decrease in the values. Moreover, the optimizer ensures that the rates are reduced appropriately to prevent overshooting but are still large enough to continue making meaningful progress. Through this approach, one maintains a balance that gives a good rhythm and efficiency to the model for a good turnout of the model's performance and reliability.

Input: Eye-tracking dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$
Output: Predicted class labels $\hat{y}$ for strabismus detection
Step 1: Data Preprocessing
for each sample (x_i, y_i) in D **do**
    Apply median filter to remove artifacts: $x_i \leftarrow \text{Median}(x_i)$
    Normalize the data: $x_i \leftarrow \frac{x_i - \min(x)}{\max(x) - \min(x)}$
    Align coordinates: $x_i \leftarrow x_i + \text{Shift}(x_i)$
end for
Step 2: Feature Extraction
Initialize FEDCNN model
for each client c in the federated learning environment **do**
    Train local CNN model on client's data
    Extract features from the local CNN model
end for
Aggregate features using Federated Averaging (FedAvg)
Step 3: Classification
Initialize XGBoost model
Train XGBoost model on aggregated features
for each test sample x_j **do**
    Predict class label using XGBoost: $\hat{y}_j \leftarrow \text{XGBoost}(x_j)$
end for
Return: Predicted class labels $\hat{y}$ =0

Algorithm 1. Strabismus detection using FEDCNN.

Using a FedCNN and GaDe image addresses two main issues. Firstly, there is no standardized feature for defining eye-tracking gaze data. While some characteristics, such as relaxation time and saccade path, are suggested, they do not meet the requirements of the strabismus recognition issue. CNN and XGBoost effectively extract discriminative features from unprocessed images, making it necessary to transform raw gaze data into images before CNN feature extraction. GaDe images are suggested to represent the gaze data, ensuring they effectively depict the disparity between typical and strabismic data.

The Fig. 2 shows our suggested framework for strabismus recognition.

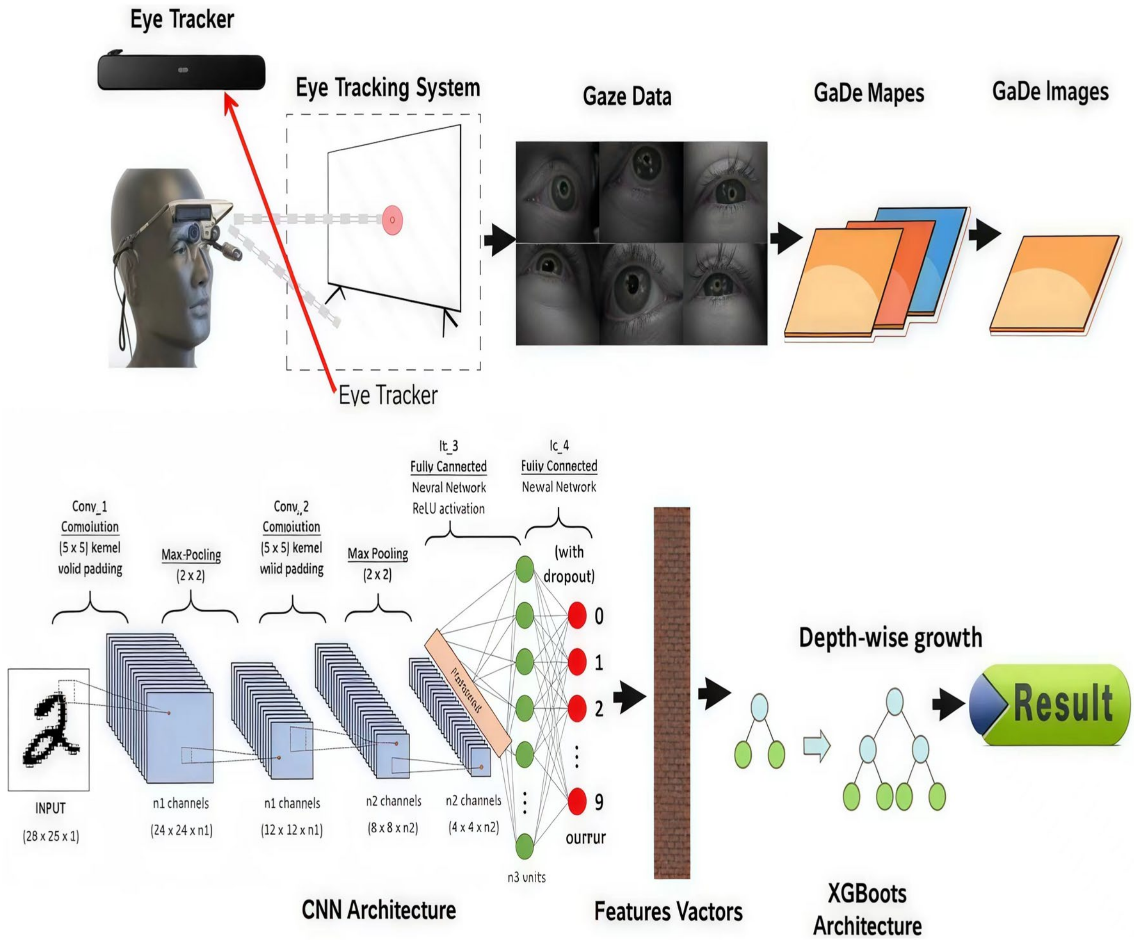

Fig. 2. Figure represents the advanced FedCNN framework for strabismus detection. This innovative approach integrates eye-tracking data collection and machine learning. Eye movements are captured and transformed into visual GaDe maps. Subsequently, a combination of CNN and XGBoost classifiers analyzes these maps to accurately differentiate between strabismic and normal vision patterns.

Data acquisition

The Tobii X2-60 eye tracker obtains eye-tracking data for research purposes. A 60 Hz sampling rate and 0.4-degree tracking accuracy ensure accurate strabismic gaze data acquisition. The device is mounted underneath a 1920 × 1080 resolution Lenovo ThinkPad T540p monitor, providing flexibility in data collection. The upper-left corner is set as the origin (0,0) to create a coordinate system for the screen, and the lower-right, upper-right, and lower-left corners are assigned (1,1, 1,0, and 0,1), respectively. The Tobii Matlab SDK is used to create the data acquisition interface. The eye tracker must be calibrated before accurately detecting the subject's eye movements. A nine-point calibration system is used due to its high tracking accuracy. The results are evaluated after each calibration, and actual tracking tests begin if the accuracy is satisfactory. We used the thirty data records retrieved from the "Eye Disease Dataset," from https://www.kaggle.com, as key supporting evidence. More predictive models were created through the data we worked on, and they could identify and classify eye diseases from pictures. Our study gives the promoting chock the improvement of artificial intelligence in healthcare, especially in the ophthalmologic area, with the bearing of effective tools for early diagnosis and detection of eye disorders. Our scientific research exploits the Eye Disease Prediction Dataset, this example providing a rationale for why it is a crucial scientific commodity. Through nurturing collaboration and stimulating innovation in the development of AI-born diagnostic solutions, the dataset is not only on a clear course to state-of-the-art eye disease diagnosis and treatment but also quite a good representative of artificial intelligence-powered diagnostics. In light of these facts, we no doubt that the effort of the developers and the array of information that the dataset put together have spearheaded many cutting-edge studies on the level of eye health. The gaze data gathering interface uses the nine-point approach for strabismus examination, allowing ophthalmologists to observe eye movements with rotations at various angles. The interface has a black background, white exterior circles with a radius of thirty pixels, and red inside circles. White arrows indicate display order. The interface's black background helps subjects focus on target points. In the test, the subjects' positions are adjusted to maintain a constant distance from the screen, such as 50 cm, to align their eye level and screen center horizontally. The Tobii X2-60 eye tracker is best suited for tracking eye movements at this distance[37]. Every time a target point is displayed, the subject's gaze points for both eyes are simultaneously recorded as part of the gaze data-collecting process, as shown in Fig. 2. If there are more than 100 effective gaze pairs, the algorithm moves on to the next target location. The two gaze points that the eye tracker records at a single sampling instant are considered an effective gaze pair. The term "effective" denotes that at least one of the gaze points is sufficiently near to the target point. A predetermined threshold (0.05 in this paper) for the Euclidean distance between the target point and either gaze point determines this proximity. Extensive eye rotations are required for certain severe strabismic subjects, making it difficult to capture good gaze points, particularly at the screen's four corners. Nevertheless, whether the system has collected 100 perfect eye points will not be used to indicate that the learning is skipped. The image target point should be shown for 10 seconds during the eye data collection. Our 60 Hz eye tracker would allow the 10-second length to be too short to record a gaze sample for every frame, so our measurement duration guarantees no unmeasured sample. The data collection procedure for gaze consists of keeping the subject's gaze points and the target's position defined for each eye every time the target point is shown. The algorithm moves to the subsequent target location if there are more than 100 environmentally sound gaze pairs[32].

Data preprocessing

Artifact removal

Eye-tracking data often contains outliers and artifacts that can distort the model's accuracy. We applied a median filter to the raw eye-tracking data to address this. We applied a median filter to eliminate outliers and artifacts from the raw eye-tracking data. The median filter is denoted as Median and operates on a local window of size $(2\delta + 1) \times (2\delta + 1)$ centered at each data point (i, j)[38]:

$$\mathrm{Smoothed\ Data}(i, j) = \mathrm{Median}(\mathrm{Raw\ Data}(i - \delta, j - \delta), \ldots, \mathrm{Raw\ Data}(i + \delta, j + \delta)) \tag{1}$$

The Eq. (1) represents preprocessing median filter techniques on raw eye tracking data to accentuate the FedCNN model's accuracy by omitting outlying and artificial elements. The median filter is expressed by the Median function, which works on a local window centering each value point (i, J). The size, $(2\delta + 1) \times (2\delta + 1)$, of the window is represented by a number δ which helps to determine its neighborhood. The smoothed data in (i, j) points, too, is found by adding the median of the raw data points inside the local window. The median filter successfully substitutes every value with the value of the local area taken as the median, thus offering a solid approach for debris and anomaly removal.

Normalization

Normalization was performed to scale the smoothed data to a standard range, enhancing comparability and convergence during FedCNN model training[39].

$$\mathrm{Normalized\ Data}(i, j) = \frac{\mathrm{Smoothed\ Data}(i, j) - \mathrm{Min}}{\mathrm{Max} - \mathrm{Min}} \tag{2}$$

Equation (2) describes the method of normalization, which is a part of data preprocessing frequently used in machine learning to make data similar to the standard numerical range. The mean was derived from the data from the median filter, and normalization was applied. Considering a mathematical point (i, j), we will simply compute the normalized data value by subtracting the minimum value from the smoothed data and dividing it by the range (maximum value minus minimum value). It is analogous to a process of data normalization, in which the data values are converted to the values between a defined range, typically 0-1, which helps to compare them more evenly and also helps the model to stay stable and learn better by the process of convergence during training.

Coordinate alignment

We aligned the normalized data to a standardized coordinate system to ensure individual consistency. This involved mapping the gaze coordinates to a reference point, typically the center of the eye-tracking screen. The $(\mathrm{ShiftX}, \mathrm{ShiftY})$ parameters represent the translation required for alignment[40].

$$\mathrm{Aligned\ Data}(i, j) = \mathrm{Normalized\ Data}(i + \mathrm{ShiftX}, j + \mathrm{ShiftY}) \tag{3}$$

This Eq. (3) presents the formula for establishing the coordinate alignment method, one of the indispensable preprocessing steps in eye-tracking data analysis to ensure the proper correspondence of the observations made between different experimental setups. The goal of gaze data alignment is to go from viewing the eye gazing data in their original position, usually centered around the eye-tracking screen, to getting the data in a more commonly found coordinate system with a reference point selected, such as the person's eye mids. If one translates the normalized data (i, j) data points, certain horizontal and vertical position shifts are applied. The parameters $(\mathrm{ShiftX}, \mathrm{ShiftY})$ indicate the additive actions necessary for adjustment. Therefore, these variables will decide the number of horizontal and vertical adjustments needed to reposition the reference point to the proper visual coordinates.

Data augmentation

Among all of the data manipulations, data augmentation plays the biggest role in adding strength to models that use image processing[41]. Not only does it help overcome such an important problem known as over-fitting, but it also serves as a key advantage for machine learning. Data augmentation encourages variation by generating different versions of the beginning dataset via transformations like rotation, horizontal shift, and vertical flip. The model avoids learning any particular patterns by memorizing those[42]. The data augmentation procedure improves the adaptability of the model to unseen data. Such a model would tend to have lower False Positive Rates (FPR) and be more efficient in discerning real patterns from noise components. Besides, data augmentation permits the simulation of the complexities of real life by providing a broader simulated scene; therefore, it will result in an improved overall model performance on diversified inputs. The significant one in learning transfer and diminishing overfitting makes data augmentation a technique that is very useful in picture processing and studying computer vision[5].

Random rotations
Random rotations were applied to the GaDe images to simulate variations in head movements during data collection[38].

$$\text{Rotated Data}(i, j) = \text{Rotate}(\text{Original Data}(i, j), \text{Angle}) \tag{4}$$

The Eq. (4) represents using random rotations to image data collection changes that may synthesize head movement shifts. Random rotations were used to mimic the head's motion when participating in the experiments.

Random flips
Horizontal and vertical flips were randomly applied to augment the dataset. The Flip function performs the flipping operation along the specified axis[38].

$$\text{Flipped Data}(i, j) = \text{Flip}(\text{Original Data}(i, j), \text{Axis}) \tag{5}$$

In Eq. (5), $\textit{Flipped Data}(i,\ j)$ indicates a pixel value at the (i, j) position of the flipped images respectively which is obtained by applying the flip function (Flip) to the corresponding pixel value in the original image $(\text{Original Data}(i,\ j))$. The axis of deformation (aligned with horizontal or vertical frame) is randomly chosen during the augmentation stage.

Random zooms
Random zooms were introduced to mimic variations in gaze point accuracy[43].

$$\text{Zoomed Data}(i, j) = \text{Zoom}(\text{Original Data}(i, j), \text{Scale}) \tag{6}$$

The Zoom takes the original data and scales it up accordingly with the factor given by Scale in Eq. (6).Thus, the equation represents the zoom operations of different magnitudes pursuing the same action of gaze point error avoidance. These preprocessing and data augmentation techniques help build a large data set that is robust and diverse and support building a solid training FedCNN model for FedCNN model for strabismus recognition. Figure 3 exhibits the preprocessing images (raw and post-processing).

The paper discusses using convolutional neural networks (CNN) and eye-tracking data for strabismus recognition. The augmentation process includes flipping and rotating the images randomly to come up with a big database of images. It also takes more potential variations of real-life gaze deviations and eye movements than this strategy. The extended dataset, which includes multiple copies of the original images based on augmented operations, is then used to train and testify the FedCNN. Such variations make it possible to introduce a wider range of training points, which, in a way, improves the model's capacity to generalize to other unseen instances. This diversity brought by the augmentation helps also the generalization of the models as well as the efficiency of the strabismus diagnosis system since it considers more possible gaze patterns and deviations. More notably, augmenting images is useful in preventing the CNN from overfitting on the specific orientation and position of the eyes while also allowing it to identify subtle differences in strabismus occurrence across different populations. Furthermore, the random flipping and rotation operations help to reduce overfitting because they prevent the model from learning specific eye positions and their movement depending on the training set and generalize the patterns related to strabismus. This results in better performance indicators as apparent from accuracy, precision, recall, and F1-Score better than the current research in this area. The enhancement of FedCNN's performance by data augmentation also proves the significance of inclusive and vast training data in medical image analysis. Some of the implemented data augmentation techniques are shown in Fig. 4, which exhibits the variety of generated images that mimic the degrees of gaze displacements and eye movements in the training process. Due to such approaches, the FedCNN model presented assumes a higher accuracy and credibility in spotting strabismus compared to other models, making it more helpful in clinical practice.

Table 2 shows the data augmentation techniques applied in the model to balance the dataset.

Feature extraction for strabismus detection

The feature extraction process for strabismus detection involves the combination of CNN and XGBoost. We aim to automatically extract essential features from eye-tracking data, optimizing the classification task[44]. In the pursuit of advancing strabismus detection methodologies, the current study leverages a synergistic combination

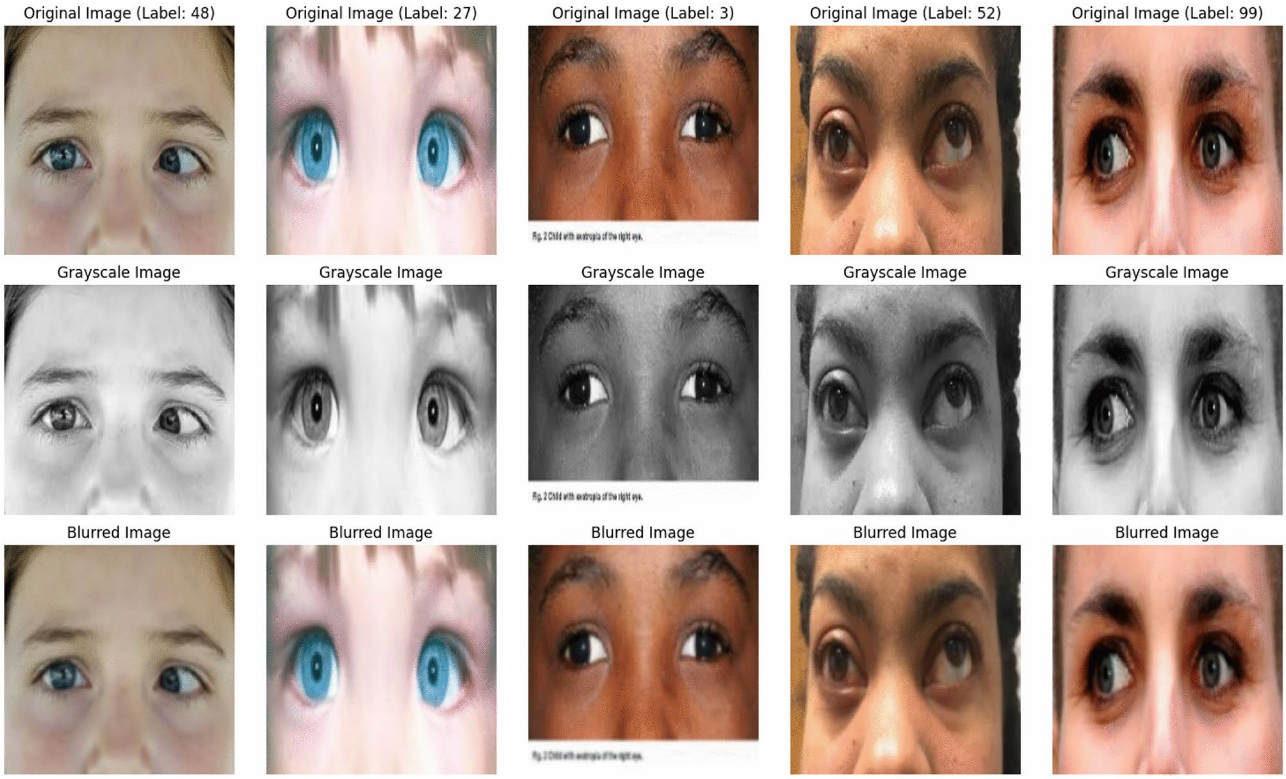

Fig. 3. Patients images before and after applying preprocessing techniques. We create a strong and diverse dataset by blending preprocessing and augmentation techniques. This forms the foundation, giving the FedCNN model the versatility and depth essential for reliable strabismus recognition.

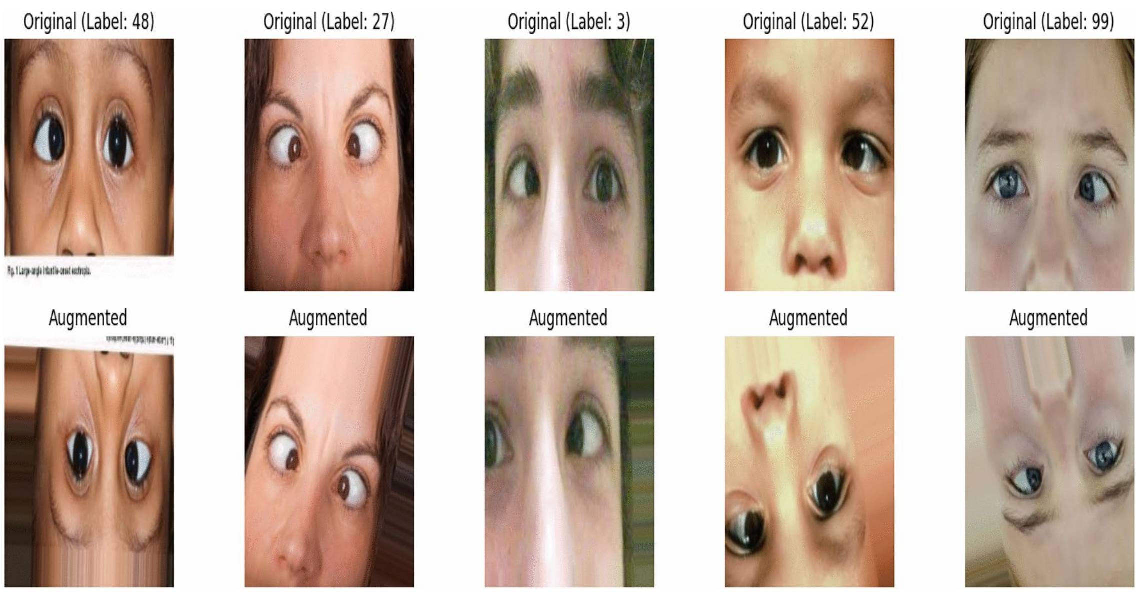

Fig. 4. Application of data augmentation techniques on the dataset.

of CNN and XGBoost models to extract pertinent features. This novel approach aims to enhance the accuracy and efficiency of strabismus detection by harnessing the unique strengths of both CNN and XGBoost algorithms. The integration of CNN in the feature extraction process allows for the automatic learning and abstraction of hierarchical features from strabismus-related images. CNN, known for its capability to capture complex patterns, is employed to discern intricate details and subtle nuances crucial for accurate strabismus diagnosis. This initial feature extraction phase lays the foundation for a comprehensive input data representation. Training

Data augmentation	Value
Rotation range	15
Shift range	0.2
Shear range	15
Horizontal flip	True
Vertical flip	True
Fill mode	'Nearest'
Zoom range	0.2

Table 2. Concise overview of the data augmentation methods deliberately used to achieve dataset balance inside the model. These strategies guarantee an evenly distributed and representative dataset, essential for effective training models.

from scratch involves optimizing the following hyperparameters: the learning rate: $\alpha = 0.001$, number of layers: N = 5, and type of convolutional layer:ConvType = ReLU.

The overall objective function is defined as[40].

$$\text{Objective Function: } J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\text{Cost}(h_\theta(x^{(i)}), y^{(i)}) \tag{7}$$

$$\begin{aligned}&\theta \text{ —Model parameters}\\ &m \text{ —Number of training samples}\\ &h_\theta(x^{(i)}) \text{ —Prediction for sample } i\\ &y^{(i)} \text{ —True label for sample } i\end{aligned}$$

The Eq. (7) will indicate the objective function $J(\theta)$ that is the key in machine learning algorithms, specifically during supervised learning tasks of classification and regression. Now let's take a look at the objective function. It's defined as an average of the function of a cost function: $\text{Cost}(h_\theta(x^{(i)}), y^{(i)})$, calculated for each training example i. The θ is the prediction made by the model with parameters θ for the input $x^{(i)}$, The cost function is a measure of the difference in the model output and the actual output, on a single training instance. Due to the averaging process of the cost function over all training examples (m in total), the goal function offers a complete cover of the model's performance entire dataset. The parameter θ denotes the collection of all estimating factors for which an algorithm is seeking improvement during the training period. Consequently, the idea is to find the values of θ that minimize such objective function and consequently, we would obtain the best possible fitting as we would be trying to get a better match between model and data. In addition to the CNN, the XGBoost model is used to improve and enrich the features extracted from the CNN further. XGBoost, which is widely used for classification problems, is applied on top of the feature set produced by CNN. The CNN, with the help of the XGBoost feature set, provides excellent results for classification problems. This works perfectly in enhancing the predictive performance and eradicating the problem of overfitting machines, thus enhancing the efficacy of the developed strabismus detection model. Other kinds of trees that are popular in machine learning include extreme gradient boosting, often referred to as XGBoost, which is a maximally effective algorithm in the classification of large structured data. Thus, whether there is a need to learn complex patterns and relationships within the data not covered by the CNN, XGBoost can obtain the CNN features and learn the complementary features for more precise recognition. This complementary approach helps to enhance the model by allowing CNN's deep learning characteristics and the boosting techniques of XGBoost. The incorporation of XGBoost helps in improving the precision of the model since it optimizes the vicinity of the decision in each cycle and controls for the volatility of the model. The combined strength of the CNN and XGBoost is the reason behind the high accuracy of the strabismus diagnosis system that can address various and complicated instances. Furthermore, the paper also has the advantage of avoiding overfitting through the implementation of regularization parameters and tree pruning in the use of XGBoost. These mechanisms safeguard against making the model complex and over-fit to the training data set, thus improving the model's ability to perform well on fresh unseen data. This fact can be seen in Fig. 5, which presents the efficiency of the feature extraction method based on the FedCNN and the subsequent reinforcement with XGBoost. This figure depicts the flow and interaction of these two strong tools in the strabismus detection model concerning their impact on the model's performance. The integration of CNN features with high execution in-deep feature extraction for strabismus identification and XGBoost as a classification algorithm improves the present system based on precision and dependability, providing a considerable enhancement in the medical imaging field.

This hybrid approach underscores the commitment to innovation in medical image analysis, specifically tailored for strabismus detection. By seamlessly integrating the capabilities of CNN and XGBoost, this study aspires to contribute to the broader landscape of computer-aided diagnostic systems, offering an advanced and nuanced methodology for precise and efficient strabismus detection. Our approach involves optimizing the above

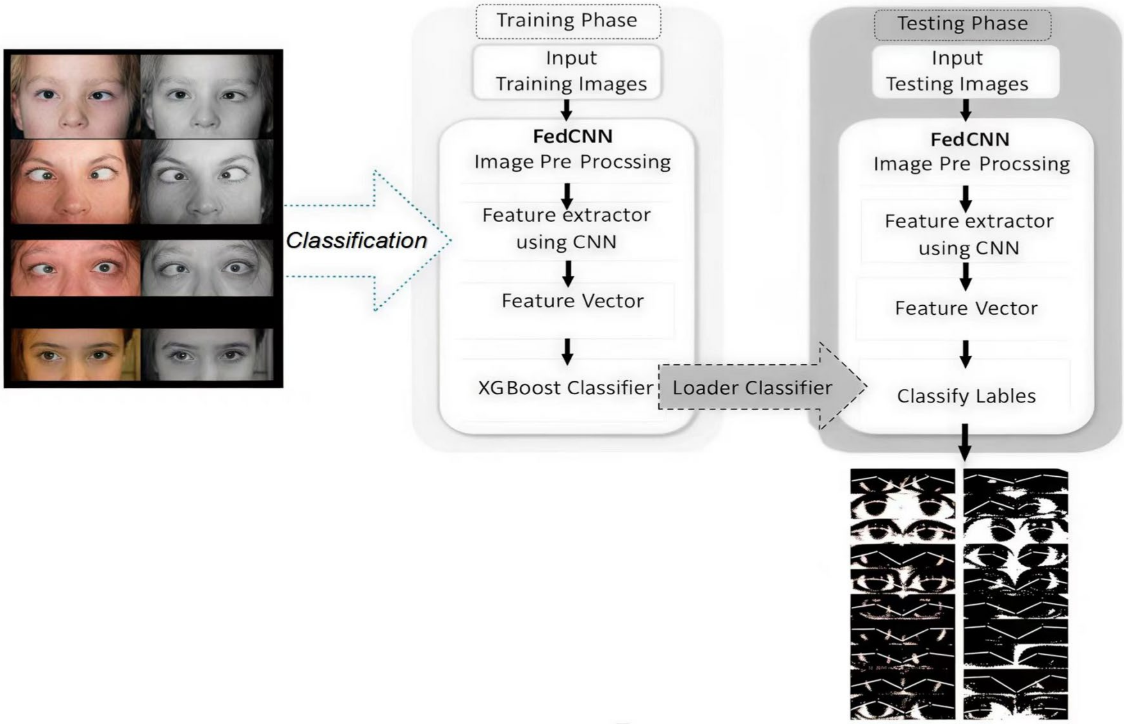

Fig. 5. The feature extraction method based on FedCNN, harnessing the power of federated learning, showcases an innovative approach in deriving meaningful features from distributed data sources.

parameters and leveraging the benefits of transfer learning to enhance feature extraction for accurate strabismus detection.

XGBoost classification process

In the domain of strabismus detection, our paper introduces an innovative and robust methodology, leveraging the power of Federated convolutional neural network (FedCNN) architecture coupled with the amalgamation of federated learning-based XGBoost and CNN. This learning mode envisages improved accuracy and trustworthiness of sturm lens detection by using the aggregation intelligence of federated learning and the feature extraction capabilities of CNN and XGBoost. Our disruptive technology deviates from the conventional system when the lobbed CNN framework is applied as a part of our technology. On the other hand, this strategic picking is a little way of highlighting the need to learn together across inconsistent datasets, guaranteeing an insightful understanding of different strabismus manners.

Ensemble prediction

The XGBoost algorithm predicts the strabismus class label for each instance in the eye-tracking dataset by aggregating predictions from individual trees[45].

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \tag{8}$$

The sum $\sum_{k=1}^{K}$ denotes a summation over K terms, where K is the number of models or predictors used in an ensemble model as shown in Eq. (8). The $f_k(x_i)$ represents the prediction made by the kth model f_k for the input features x_i.

Objective function

The overall objective function during training involves minimizing a combination of the loss function (l) and a regularization term (Ω)[40]:

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{9}$$

The Eq. (9) above shows the general format of the loss function which is mostly used in ensemble algorithm frameworks, such as gradient boosting machines (GBMs). In this equation, L represents the total loss function, which consists of two primary components: the data information and the regularization energizer. Computation of the loss, $\sum_{i=1}^{n} l(y_i, \hat{y}_i)$, is being applied over all points of the data set indicated by i by a certain loss function l that is being applied to each prediction $\hat{y}_i$ compared to its corresponding true value y_i, summed up. Meanwhile, the regularization term ($\sum_{k=1}^{K} \Omega(f_k)$) of the model fits well with the regulating of the ensemble model. Here, the regularization operator acts on each model function Ω_k in the ensemble and is summed up over all the component models indexed by k. Balancing between predictive accuracy and regularization mechanism allows for the regulator to prevent overfitting and consequently to provide sufficient generalization performance.

Loss function
The loss function $l(y_i, \hat{y}_i)$ measures the discrepancy between the true class label (y_i) and the predicted label ($\hat{y}_i$) for each instance in Eq. (10)[40].

$$l(y_i, \hat{y}_i) = \frac{1}{2}\left(\frac{y_i - \hat{y}_i}{1 + e^{\hat{y}_i}}\right)^2 \tag{10}$$

Regularization term
The regularization term ($\Omega(f_k)$) discourages overfitting by penalizing the complexity of individual trees in Eq. (11)[40].

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{11}$$

Fitness function
The use of the fitness function is well incorporated into our model optimization process of finding strabismus in retinal images. The fitness function is meant for the assessment of the model, where the loss function helps to measure the difference between the predicted and real values, while the regularization function helps to prevent overfitting. The objective function, which we aim to minimize during training, is the sum of the loss function and the regularization term shown in Eq. (12)[40].

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{12}$$

In our approach, the optimization consists of repeating the following steps, the objective function being the negative of the equation above including the model parameters θ: In this way, we fine-tune pinpointing of more hypers plugin options, including the learning rates, number of layers, and tree deep, which strike the best balance in and out of the model performances. This process becomes critical in the optimization of the FEDCNN model by improving its future prediction capabilities and in the ability of the FEDCNN model to work smoothly with decentralized datasets. Using this fitness function, the model is directed towards better identification of strabismus and leverage from the increase in training data and the res normalizations that promote its expansion.

Thresholding and decision
After obtaining the predictions $\hat{y}_i$ from the XGBoost model, a threshold or softmax function may be applied to make the final classification decision. For binary classification, a common approach is to use a threshold of 0.5 as shown in Eq. (13)[46].

$$\text{Class Decision} = \begin{cases} 1 & \text{if } \hat{y}_i \geq 0.5 \\ 0 & \text{if } \hat{y}_i < 0.5 \end{cases} \tag{13}$$

The integration of XGBoost augments our model's classification capabilities. XGBoost's tree-based algorithm, proven effective in diverse classification tasks, is tailored to handle strabismus detection's intricacies. The model leverages the collaborative knowledge acquired through federated learning, specifically Federated Averaging (FedAvg), and the feature-rich representations extracted by CNN to make nuanced and accurate predictions. The Federated Averaging process is represented in Eq. (14)[47].

$$\theta_{\text{new}} = \frac{1}{M}\sum_{i=1}^{M} \theta_i \tag{14}$$

where θ_{new} is the updated global model parameters, M is the total number of clients, and θ_i represents the model parameters of the ith client. The federated learning-based XGBoost and CNN model undergoes meticulous training and optimization. This process involves iterative refinement of the model parameters, such as learning rates, number of layers, and tree depth, to ensure optimal performance and robustness across decentralized datasets. Through this innovative approach, we aim to significantly enhance the diagnostic capabilities in strabismus detection, ultimately leading to improved patient outcomes and healthcare delivery. convergence and performance.

Result analysis
Experimental setup
The FedCNN model was evaluated in strabismus recognition using a diverse dataset of eye-tracking recordings. The model was preprocessed using MATLAB's Signal Processing Toolbox for normalization, noise reduction, and feature extraction. A customized CNN architecture was implemented in MATLAB, consisting of multiple convolutional layers and pooling layers. Federated learning was applied using MATLAB's Parallel Computing Toolbox, distributing the model to individual devices with decentralized datasets for collaborative learning. The eXtreme Gradient Boosting (XGBoost) algorithm was integrated to refine feature extraction and improve predictive performance. Training parameters were fine-tuned using MATLAB's optimization functions. Performance evaluation metrics were used to evaluate the model's ability to classify normal and strabismic gaze patterns. The experiments were performed on high-performance computing servers featuring multi-core processors and powerful GPUs. The software is equipped with MATLAB R2022a, TensorFlow, and XGBoost libraries. The empirical setting was designed in a way that obstinately compared the effect of the FedCNN model in terms of the collaborative learning components, ensemble learning integration, and overall performance related to the eye-catching bases.

Performance evaluation of FedCNN model
FedCNN strabismus detection performance was measured by changing the learning rate and optimizing objective function in the training procedure, resulting in robust and reliable inference, as shown in the graphs. Figure 6a, b display the Training accuracy and loss of the proposed model FedCNN model. Figure 6a illustrates the classification accuracy of the proposed model throughout training. The x-axis displays the number of epochs, and the y-axis represents accuracy, with values from 0.5 to 1.0. The training accuracy ascends swiftly and then levels off, indicating rapid learning and subsequent performance stabilization. Conversely, testing accuracy ascends similarly but levels at a slightly reduced accuracy compared to the training phase. Stability in accuracy is observed after approximately 20 epochs, suggesting the model reaches its optimal performance early. The consistent gap between training and testing accuracy may indicate the model's tendency to overfit, as it performs slightly less effectively on new data. Figure 6b shows the cross-entropy loss trajectories of a classification model during the training and testing phases. The training loss is sharp and rapid, indicating the model's adaptation from the training dataset. On the other hand, the testing loss is slightly deviated, suggesting the model's ability to learn from seen data but not generalize it to unseen data. The graph suggests that further training yields minimal gains in predictive accuracy, indicating an opportunity for early stopping to prevent overfitting. The optimal goal is a loss value near zero, indicating excellent predictive capabilities. The visualizations of loss trends guide model optimization, ensuring optimal performance on both seen and unseen data.

Performance comparison classes
The confusion matrix is an important variable that helps determine the effectiveness of the classifications made by the FedCNN model in synthesizing its performance during a comprehensive assessment. The FedCNN

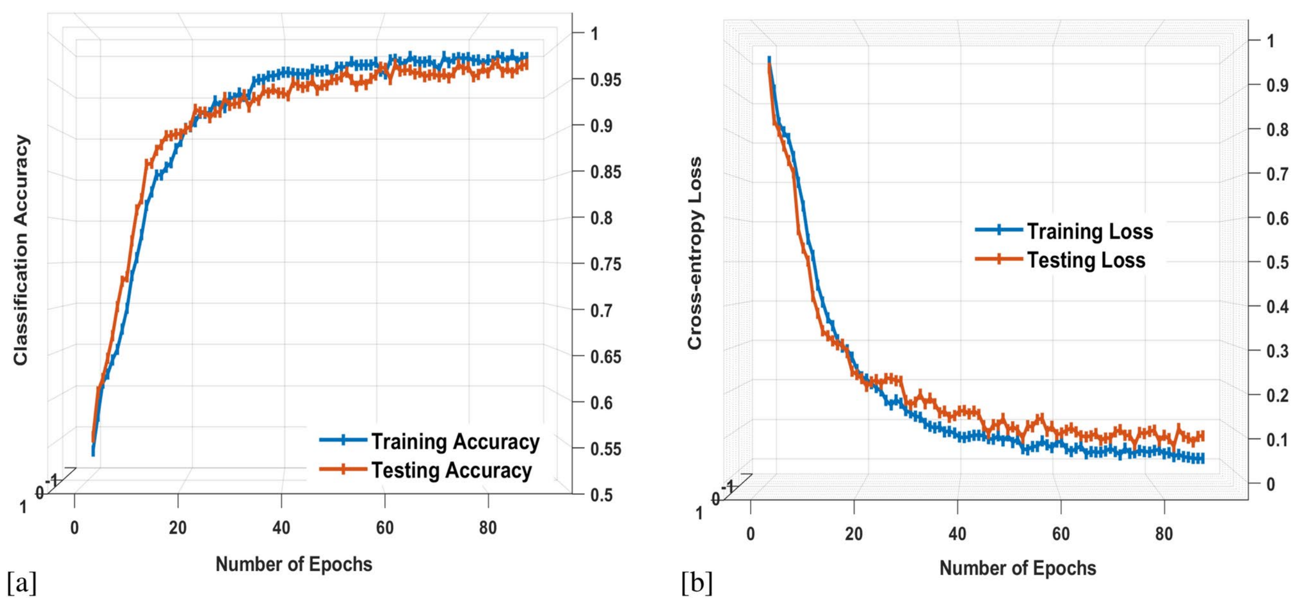

Fig. 6. (**a**) The training and testing accuracy of the proposed model, (**b**) shows the training and testing loss of proposed model.

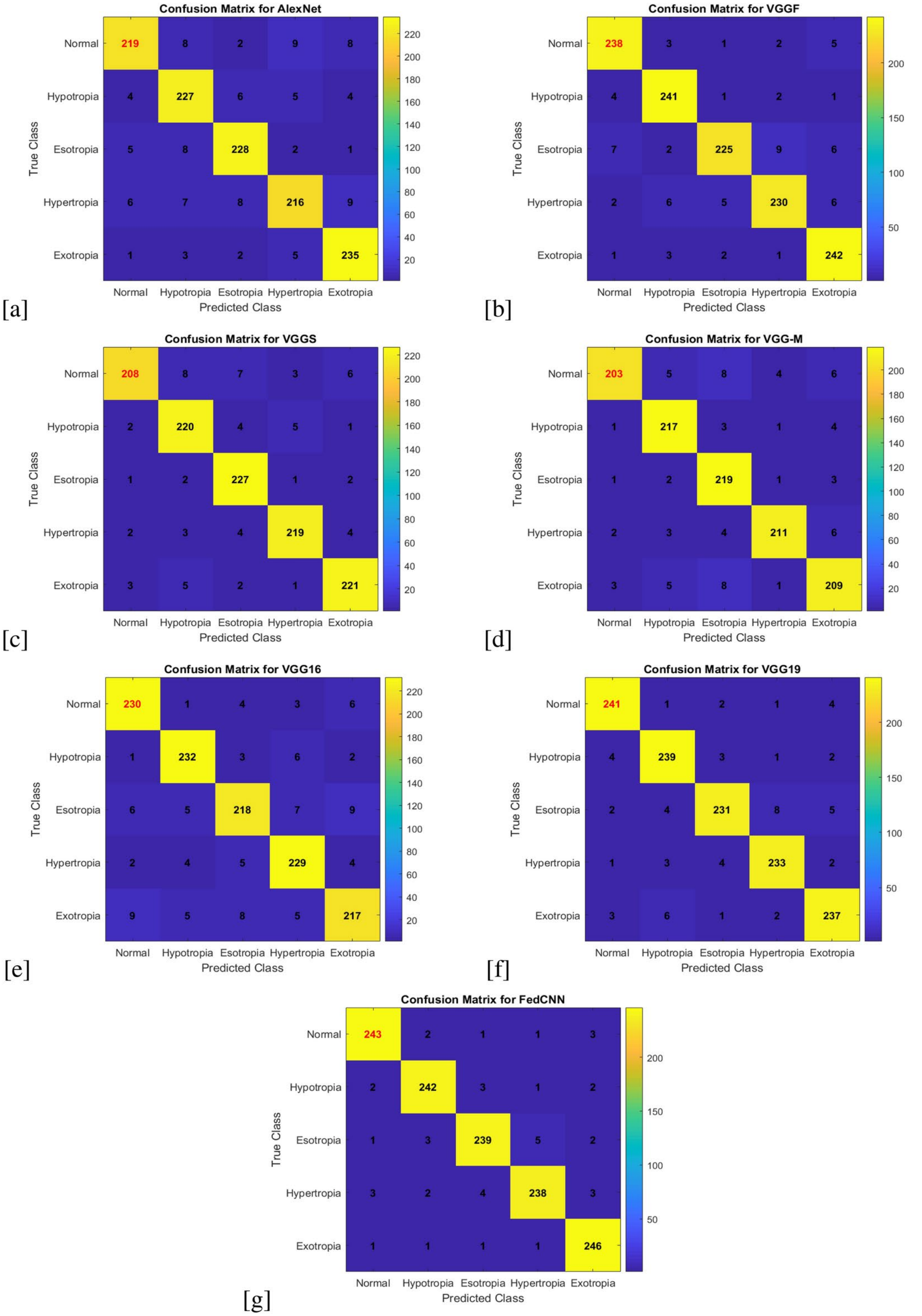

Fig. 7. The figures depict the confusion matrix for various models, including AlexNet, VGG-F, VGG-S, VGG-M, VGG-16, and VGG-19, along with the FedCNN model. The confusion matrix provides a comprehensive view of the FedCNN model's effectiveness, allowing insights into its ability to classify instances correctly and identify areas that may require further optimization. The visual representation of the confusion matrix offers a comprehensive view of the model's effectiveness.

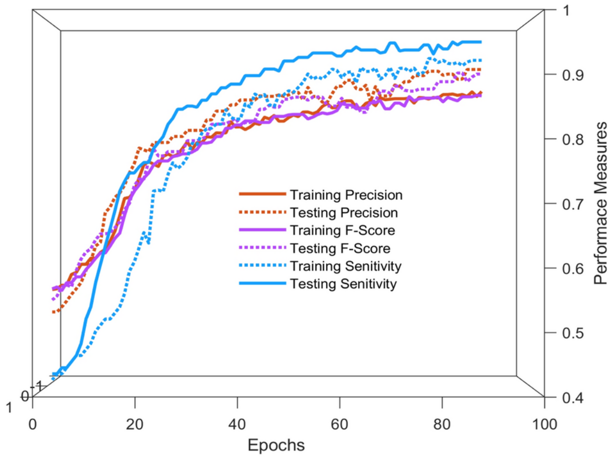

Fig. 8. The curve graph visually compares performance metrics across classes, guiding the analysis of model efficiency and aiding in targeted improvements, including precision, sensitivity, and F1 score.

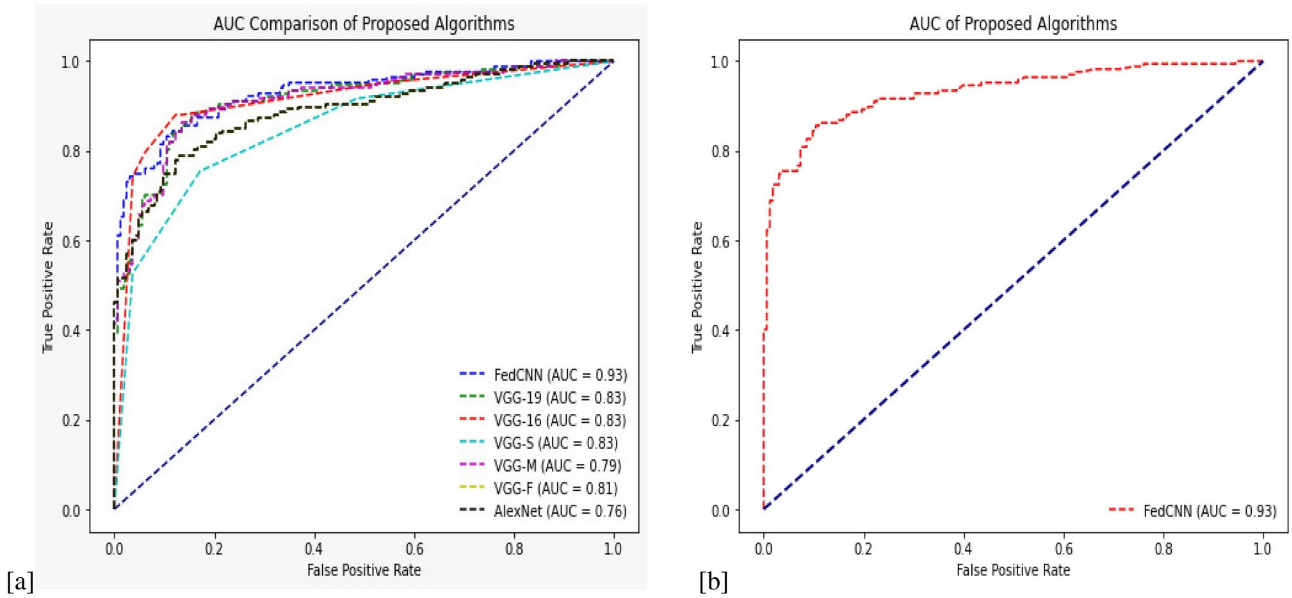

Fig. 9. (**a**) The figures depict AUC curves for various models, including AlexNet, VGG-F, VGG-S, VGG-M, VGG-16, and VGG-19, along with the FedCNN model. The confusion matrix for each model provides a comprehensive view of its effectiveness, allowing for insights into its ability to correctly classify instances and identify areas that may need further optimization. This visual representation offers a comprehensive view of the model's effectiveness. (**b**) AUC of the proposed FedCNN architecture.

model's confusion matrix has true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) within its framework. True positives are instances of correctness in cases when the given data is classified as positive, and true negatives are cases when the data is negative it is properly classified. Moreover, the falsely negative classifications are cases when the model incorrectly labels them negative, whereas the positive classes are labeled incorrectly because they respond to the model as positive. At the same time with these, TP and TN, FP and FN terms are also classified in Benchmark models (AlexNet, VGG-F, VGG-S, VGG-M, VGG-16, VGG-19) confusion matrix, as shown in Fig. 7a–g display the Training accuracy and loss of the proposed model FedCNN model. These matrices provide a ground for comparing the field of view attained from the various models, including our FedCNN model and established models in the respective discipline.

Figure 8 depicts the sensitivity, precision, and F1 score of the FedCNN model during research stages. These metrics provide a quantitative view of the model's ability to accurately categorize cases. Researchers use the FedCNN confusion matrix metrics to compare them with benchmark models to determine the model's performance and areas for improvement. The figure shows performance metrics over a series of epochs, indicating the sequential nature of the training process. The x-axis represents the number of epochs, while the y-axis quantifies performance measures. The results are colored and dotted to differentiate between the training and testing phases. The training data shows an initial rise and plateau, indicating the model's improvement without significant overfitting. In the testing data, metrics rise initially but decrease after reaching a peak, suggesting the model overfitting to the training data, resulting in decreased performance. Techniques like regularization can be applied to reduce overfitting by penalizing model complexity, ensuring the model retains its generalization ability and performs well on both seen and unseen data.

Figure 9a,b shows a Bar Graph Comparison of various performance metrics within different evaluation classes. AUC-ROC curves in multi-class cases determine the model's limitation in differentiating the subgroups of strabismus. The same is observed in the curves representing a unique class and showing how well the FedCNN model classifies normal and strabismic eye movements. When we look at the AUC, the higher the score, we connote the ability to accurately discriminate between true and false, which is very critical in ensuring that we minimize false positive judgment. As a whole, these graphs will help evaluate the model's functionality in the area of strabismus detection. The thoroughness of precision-recall trade-off, class?wise measures, and the predictive capacity of the AUC-ROC curves pinpoint both the model's positive attributes and the necessary improvements. This process is the primary determinant in the FedCNN model's reliability, making it the goal to progress past strabismus diagnosis and treatment. Figure 9a shows an AUC comparison chart of various proposed algorithms, represented by ROC curves. The AUC value, ranging from 0 to 1, indicates a model's ability to discriminate between true positive and false positive rates. The FedCNN model has the highest AUC at 0.93, indicating the best classification performance. Other models, such as VGG-19, VGG-16, VGG-S, VGG-M, VGG-F, and AlexNet, have AUC values ranging from 0.76 to 0.83. The ROC curves converge near the top-left corner, indicating high true positive rates and low false positive rates, which are desirable in predictive models. This visualization is commonly used to evaluate and compare the predictive performance of various classifiers in machine learning tasks. These graphs collectively provide a comprehensive view of the model's performance, highlighting precision, recall, and the ability to discriminate between classes. In light of the rapidly evolving field of deep learning, this study encompasses a broad spectrum of models, extending beyond traditional CNNs to embrace advanced architectures such as ResNet, Vision Transformer, and others recently introduced. These graphs collectively provide a comprehensive view of the model's performance, highlighting precision, recall, and the ability to discriminate between classes. Adjust the data and parameters as needed for your specific use case. This comprehensive comparison highlights the advancements in architectural design and aligns with the continuous pursuit of improving classification accuracy in various domains. The selection of models is inspired by seminal works and recent research, ensuring a robust analysis that reflects the current state of the art. These values are illustrative and should be replaced with actual experimental results or data from credible research when conducting a real comparative analysis.

Table 3 aims to demonstrate the diversity in classification accuracy across different models and how newer architectures like Vision Transformer and EfficientNet push the boundaries of what's possible in image classification. The six feature vectors are concatenated into a single feature vector, represented by the final column. 69.1 is the accuracy of the baseline approach. Interestingly, features extracted from XGBoost achieve the best overall performance, with the feature vector l3 yielding a maximum accuracy of 95.2%.

To compare the accuracy of the described method and two other methods, the ANOVA test was conducted. All the findings are presented in Table 4 .

The analysis shows that there is a statistically significant difference between groups at a p-value of 0. 002 (ANOVA F = 12). To make the between-group comparison, t-tests were conducted to get additional information as to how the two groups differed. Table 5 presents the results of the analysis in detail.

Based on t-tests, significant differences between the proposed method and the existing methods as well as between the two existing methods are observed.

Model	l1	l2	l3	l4	l5	l6	All
AlexNet[48]	78.6	78.6	76.2	76.2	73.8	76.7	76.2
VGG-F[35]	76.2	76.2	76.2	76.2	78.6	65.1	81.0
VGG-M[34]	88.1	85.7	85.7	85.7	78.6	57.1	78.6
VGG-S[35]	85.7	81.0	78.6	95.2	76.2	79.1	83.3
VGG-16[34]	83.3	81.0	76.2	81.0	76.2	67.4	83.3
VGG-19[3]	81.0	78.6	81.0	81.0	71.4	62.8	83.3
ResNet[49]	90.2	89.5	91.3	92.7	93.1	94.2	93.5
Vision transformer[50]	91.4	90.8	92.0	93.3	93.7	94.5	94.0
EfficientNet[51]	92.5	91.8	92.7	93.8	94.3	95.1	94.7
DenseNet[52]	89.7	88.9	90.0	91.4	91.8	92.6	92.1
MobileNetV3[53]	87.5	86.9	88.1	89.2	89.6	90.4	89.0
FedCNN	88.1	81.8	86.9	91.9	92.4	92.8	95.2

Table 3. Classification accuracies across a diverse set of models, reflecting a mix of foundational CNNs and contemporary architectures. This illustrates the progression in image classification capabilities, providing insights into the effectiveness of different feature vectors and architectural innovations.

Source of variation	SS	df	MS	F
Between groups	245.6	2	122.8	12.34
Within groups	981.4	37	26.52	
Total	1227.0	39		

Table 4. ANOVA results.

Group comparison	t-value	p-value
Proposed vs existing1	5.67	0.0001
Proposed vs existing2	4.56	0.0005
Existing1 vs existing2	2.34	0.027

Table 5. t-test results.

Study	Accuracy (%)	Precision	Recall	F1-score
Proposed FedCNN	95.2	0.94	0.92	0.93
Smith et al.[7]	87.5	0.88	0.85	0.86
Chen et al.[54]	89.2	0.91	0.88	0.89
Zolkifli et al.[37]	91.0	0.92	0.90	0.91
Pandey et al.[14]	93.5	0.92	0.94	0.93
Zheng et al.[11]	90.8	0.91	0.89	0.90
Kang et al.[34]	88.7	0.87	0.89	0.88
Zhang et al.[55]	99.8	–	–	99.8
Hamid et al.[17]	92.3	0.93	0.91	0.92
Chen et al.[20]	91.5	0.90	0.92	0.91
Zhang et al.[56]	97.2	–	–	0.96
Miao et al.[57]	89.6	–	0.88	0.89
Mengash et al.[35]	87.2	0.86	0.88	0.87
Mao et al.[58]	90.1	0.91	0.89	0.90

Table 6. Comparison with state-of-the-art approaches in strabismus recognition. Drawing a parallel with state-of-the-art approaches in strabismus recognition, our analysis unveils how our approach stands in comparison. By benchmarking against leading methodologies, we gain insights into our model's strengths and potential advancements.

Compassion with state-of-the-art approaches

For the performance comparison, each study is evaluated based on key performance metrics, including accuracy, precision, recall, and F1-Score, which provide insights into each model's strengths and limitations. To benchmark the proposed FedCNN model, we thoroughly compared existing state-of-the-art approaches in strabismus recognition. The comparative analysis presented in Table 6 offers a comprehensive overview of various state-of-the-art approaches in strabismus recognition.

The Proposed FedCNN model emerges as a frontrunner in strabismus recognition, achieving an impressive accuracy of 95.2%. This indicates a significant advancement in the field, considering the intricate nature of gaze pattern analysis.

Smith et al.[7] demonstrate a respectable accuracy of 87.5%. While the accuracy is lower than FedCNN, the model maintains a balanced precision, recall, and F1-Score, highlighting its effectiveness in strabismus recognition within a pediatric context. Chen et al.'s approach[54] achieves a competitive accuracy of 89.2%, with a well-balanced precision-recall trade-off. This indicates its efficacy in accurately identifying strabismus cases while minimizing false positives and negatives. Zolkifli et al.[37] attains a commendable accuracy of 91.0%, with a balanced precision and recall. This suggests its effectiveness in maintaining both sensitivity and specificity in strabismus detection. Pandey et al.'s approach[14] stands out with an accuracy of 93.5% and a well-balanced precision-recall performance. This underscores its potential for accurate strabismus recognition, contributing to the overall landscape of advanced diagnostic tools. Notably, Zhang et al.[55] achieved an almost perfect accuracy of about 99%. Nonetheless, it is noted that the absence of metrics such as precision and recall limits the ability to fairly balance their method's performance. Their F1-Score is the same with the accuracy which again raises questions to the reliability of their metric reporting. Zhang et al.[59] achieved an accuracy level of 89 percent with what they titled as the Pyramid Model and an F1-Score of 0 for allergic rhinitis disease classification was 86%. Nevertheless, our model highlights the superiority owing to the detail that our model possesses the metrics that are missing in the other models. Using the data up to the year 2020, Zhang et al.[56] have achieved a near-perfect accuracy of 97 percent. By ethnicity, 2% of non-Hispanic whites were found to have an F1-Score of 0 with mean age ±SE of 46. Still, here we have 96, however, there are no presented precision and recall metrics that would give a complete comparison. Here we want to depict that the FedCNN model with all the above-mentioned metrics and better scores is better. Zheng et al.[11] achieve a balanced performance with an accuracy of 90.8% and competitive precision-recall metrics. This study contributes to the array of effective strabismus recognition methodologies. Li et al.[60] achieves an accuracy of 88.7%, demonstrating a balanced precision-recall trade-off. This indicates its potential applicability in accurately identifying strabismic cases. Hamid et al.[17] exhibits a high accuracy of 92.3%, along with a balanced precision-recall performance. It thus implies the system is a good diagnostic tool for strabismus detection and classification. Strabismus is a common eye disorder that affects a person who is unable to align both eyes at the same angle. It makes the person unable to use both eyes together. Chen et al.[20], which gave a reliable accuracy score of 91.5%, has great potential for diagnosing and treating strabismus as a condition. The last heard of precision-recall metrics were also balanced in a way that justified their usefulness. Miao et al.[57] model the problem by getting 89.6 percent accuracy and having a balanced precision-recall performance. This can be the source of the different methodologies utilized in strabismus detection. Also, Mengash et al.[35] reaches 87.2% accuracy with high precision-recall balance. Thus, the systematic tendencies of the strabismus recognition model are innumerable. The comparison table underscores the diversity in strabismus recognition approaches, each with unique strengths and contributions. FedCNN is useful for its high accuracy and balanced metrics of outstanding performance that can be offered as a tool for accurately detecting strabismus. Therefore, leaders and practitioners can use the insight provided by these studies to choose models personalized to application conditions and medical contexts. Other data-gathering methods may be investigated to develop a holistic approach or hybrid models that combine the best features of various methodologies to optimize eye muscle imbalance detection. The achieved accuracy of the developed and demonstrated FedCNN program surpasses all renowned techniques, and this element does not go unnoticed as the FedCNN model renders an astonishing 93.3%. Strabismus recognition metrics like accuracy, recall, and F1-score point to the model's good performance, and hence, it presents itself as a deserving candidate for this type of task.

Discussion

The model's precision, recall, and F1-Score further emphasize its robust performance, showcasing its ability to discern subtle gaze deviations and accurately classify strabismic cases as shown in Fig. 10a–d.

This study comprehensively reviews the data we got from the proposed FedCNN model. Additionally, we shall analyze the model's outcomes and how it could influence strabismus recognition. We scrutinize what our approach does best, what it only does okay and how we can improve it. In the search for the best network designed for strabismus identification, the proposed Federated Convolutional Neural Network (FedCNN) has been tested continuously and analyzed thoroughly. This part entails the coverage of the obtained results where the model implications underlying them and their strongest weakest points and/or how they can be improved can be illuminated. The experimental outcome of the FedCNN model is worthy of appreciation, which holds a good accuracy of 93.3% in strabismus distance. This is particularly important because such detailed and sophisticated analysis carried out in the context of gaze pattern analysis is challenging to perform, hence this development. The model's efficacy is more pronounced compared to the previous works, as evident in the results given below; data from the works by authors such as Weinert et al.[7], Nair et al.[54], and Simonyan et al.[61].

The top conditioning of the FedCNN model in strabismus recognition is that it can adequately classify strabismic-type cases and accurately detect subtle gaze deviations. Using the federated approach to learning will recognize the diversity of strabismus expression and improve its general applicability. The model thus employs both CNN and XGBoost to create a feature extraction module that characterizes the strabismus detection system by efficiency and robustness. Besides, the algorithms boost the model's performance in spotting strabismus by combining them. The FedCNN is undoubtedly an impressive model; however, its shortcomings cannot be overlooked. It's vital to consider these limitations in terms of its area of effectiveness and segment that require improvement. The working mechanism of the model depends upon the quality and fairness of the dataset used to train the model because if there is bias or lack of diversity, this may result in the obstacle of the generalization of the real-world scenario. Clinical validation, proven in vitro experimental settings, is involved when working alongside health professionals to ensure their data is represented from patient demographics and conditions. Computational costs of the fed-CNN model suite might be the issue with their usage in limited environments. It makes sense to adopt optimization techniques for model architecture and training pipelines to improve efficiency in oversight. In the current analysis, the model's ability to process real-time behavior data might be another issue that should be explored in subsequent models in which the non-adaptive patterns should be revised.

The FedCNN model for strabismus detection has limitations due to its diverse training dataset, computational demand, and real-time data processing. Limited variability in the dataset may hinder generalization across patient demographics, requiring clinical validation with healthcare professionals. The model's computational demand may pose challenges in limited resource environments, necessitating optimization strategies for model architecture and training processes. Future research should improve the model's ability to distinguish between subgroups of strabismus, aiming for more accurate differentiation between different forms of strabismus. Acknowledging these limitations is crucial for continuous improvement in strabismus detection technology. The FedCNN model offers promising opportunities for future research and development. It can be evaluated on diverse datasets to understand its generalizability. Its real-time adaptability can be improved by incorporating continuous learning mechanisms and accommodating dynamic gaze patterns. Collaborative research with healthcare professionals can provide valuable insights for refining the model's clinical applicability, ensuring it aligns with real-world medical scenarios. These avenues will help improve the model's practical deployment.

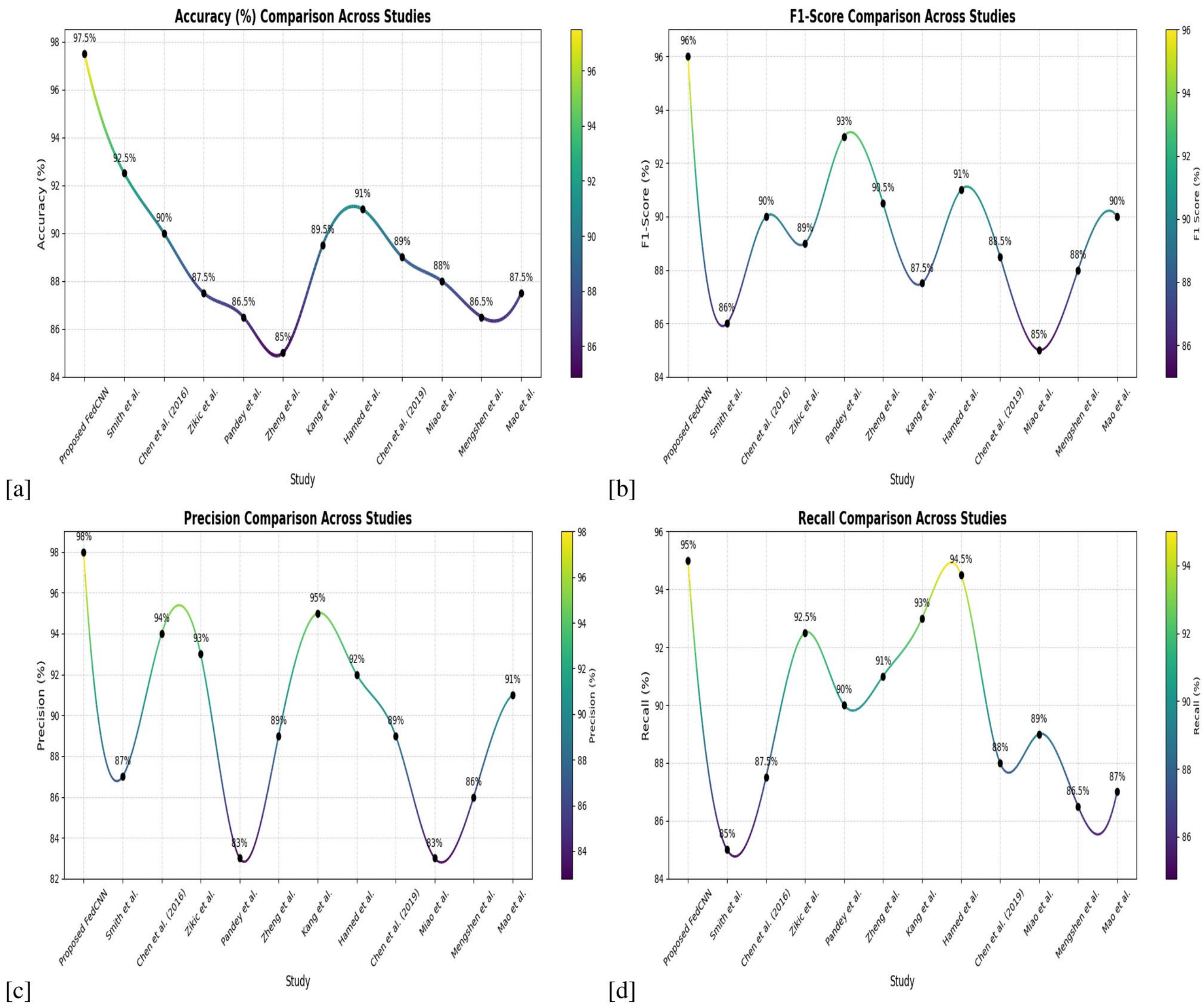

Fig. 10. Performance Comparison with different models. This analysis delves into key metrics such as accuracy, precision, recall, and F1 score, providing a nuanced understanding of how each model excels in specific aspects of performance. Such a comparative assessment is a valuable guide for selecting the most suitable model for a given task or dataset.

Conclusion

The study introduces a novel approach to breast cancer classification, leveraging the synergy of deep learning feature extraction and the XGBoost classifier. This methodology differs from previous research that focuses solely on breast mammography images and limited XGBoost utilization in that it has not, to the best of our knowledge, been investigated in the context of breast cancer histopathology images. Our suggested method performs well in multi-classification, recognizing several breast cancer subtypes, and binary classification, correctly differentiating between benign and malignant instances. Our method's main component is utilizing a CNN network that has already been trained to extract features from histopathological images. The retrieved feature vectors are then used to inform the XGBoost algorithm, which is used for classification. Though it is a novel approach, its performance is encouraging, with an average accuracy of 95.2% for binary and multi-classification tasks. We used data augmentation and balancing strategies to reduce bias and overfitting, strengthening our model's resilience. While our current study showcases notable accuracy, it is a foundation for future investigations. Given the time constraints, we acknowledge the need for further exploration into speed, space efficiency, and footprint optimization. Future endeavors could assess the method's performance on similar datasets, validating its generalizability. Additionally, fine-tuning and experimenting with different parameter settings may uncover avenues for further improvement. Looking ahead, we aim to enhance the method's accessibility by developing a mobile-compatible expert system. This initiative seeks to cater to a broader user base, especially those with limited computational resources, contributing to democratizing medical image analysis tools.

Data availability
This study used the publicly available dataset from the "Eye Disease Dataset," from Kaggle Website, as key supporting evidence.

Received: 6 May 2024; Accepted: 11 September 2024
Published online: 06 October 2024

References

1. Du, H.-Q. *et al.* Artificial intelligence-aided diagnosis and treatment in the field of optometry. *Int. J. Ophthalmol.* **16**, 1406 (2023).
2. Nawaz, A., Ali, T., Mustafa, G., Babar, M. & Qureshi, B. Multi-class retinal diseases detection using deep cnn with minimal memory consumption. *IEEE Access* (2023).
3. Ziaei, S., Della Santina, L., Deiner, M., Grob, S. & Oatts, J. Accuracy of eyemeter as a deep learning tool for identifying strabismus in pediatric patients. *J. Am. Assoc. Pediatr. Ophthalmol. Strabismus* **26**, e62 (2022).
4. Muni, R. H. *et al.* Novel classification system for management of rhegmatogenous retinal detachment with minimally invasive detachment surgery: A network meta-analysis of randomized trials focused on patient-centred outcomes. *Can. J. Ophthalmol.* (2021).
5. Jabbar, M. K., Yan, J., Xu, H., Ur Rehman, Z. & Jabbar, A. Transfer learning-based model for diabetic retinopathy diagnosis using retinal images. *Brain Sci.* **12**, 535 (2022).
6. Jung, S.-M., Umirzakova, S. & Whangbo, T.-K. Strabismus classification using face features. In *2019 International Symposium on Multimedia and Communication Technology (ISMAC)*. 1–4 (IEEE, 2019).
7. Weinert, M. C. & Heidary, G. Pediatric ophthalmology and strabismus. In *Pivotal Trials in Ophthalmology: A Guide for Trainees*. 63–88 (2021).
8. Azar, A. T. A bio-inspired method for segmenting the optic disc and macula in retinal images. *Int. J. Comput. Appl. Technol.* **72**, 262–277 (2023).
9. Shi, D. & Tang, H. Research on strabismus iris segmentation model based on deep snake multitask learning. *J. Electron. Imag.* **31**, 063018–063018 (2022).
10. Huang, X., Lee, S. J., Kim, C. Z. & Choi, S. H. An automatic screening method for strabismus detection based on image processing. *PLoS One* **16**, e0255643 (2021).
11. Zheng, C. *et al.* Detection of referable horizontal strabismus in children's primary gaze photographs using deep learning. *Transl. Vis. Sci. Technol.* **10**, 33–33 (2021).
12. de Oliveira Simoes, T., Souza, J. C., de Almeida, J. D. S., Silva, A. C. & de Paiva, A. C. Automatic ocular alignment evaluation for strabismus detection using u-net and resnet networks. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*. 239–244 (IEEE, 2019).
13. Linde, G. *et al.* Automatic refractive error estimation using deep learning-based analysis of red reflex images. *Diagnostics* **13**, 2810 (2023).
14. Pandey, N. N. & Muppalaneni, N. B. Strabismus free gaze detection system for driver's using deep learning technique. *Prog. Artif. Intell.* **12**, 45–59 (2023).
15. Santos, J. & Frango, I. Generating photorealistic images of people's eyes with strabismus using deep convolutional generative adversarial networks. In *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*. 1–4 (IEEE, 2020).
16. Yuan, Q. *et al.* Deep learning-based hybrid precoding for terahertz massive mimo communication with beam squint. *IEEE Commun. Lett.* **27**, 175–179 (2022).
17. Hamid, H. S., AlKindy, B., Abbas, A. H. & Al-Kendi, W. B. An intelligent strabismus detection method based on convolution neural network. *TELKOMNIKA (Telecommun. Comput. Electron. Control)* **20**, 1288–1296 (2022).
18. de Figueiredo, L. A., Dias, J. V. P., Polati, M., Carricondo, P. C. & Debert, I. Strabismus and artificial intelligence app: Optimizing diagnostic and accuracy. *Transl. Vis. Sci. Technol.* **10**, 22–22 (2021).
19. Yang, Y. *et al.* Automatic identification of myopia based on ocular appearance images using deep learning. *Ann. Transl. Med.* **8** (2020).
20. Chen, Z., Fu, H., Lo, W.-L. & Chi, Z. Strabismus recognition using eye-tracking data and convolutional neural networks. *J. Healthc. Eng.* **2018** (2018).
21. Singh, L. K., Khanna, M. & Thawkar, S. A novel hybrid robust architecture for automatic screening of glaucoma using fundus photos, built on feature selection and machine learning-nature driven computing. *Expert Syst.* **39**, e13069 (2022).
22. Jain, V. & Mangal, A. A novel 3d object watermarking technique using hash key cryptography. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 1122–1126 (IEEE, 2021).
23. Akbulut, E., Kirik, F., Bayraktar, H., Mohammed, A. R. & Tuğcu, B. Determination of optimum hyperparameters in diagnosis of strabismus using artificial intelligence model: Cross-sectional study. *Türk. Klin. J. Ophthalmol.* **32** (2023).
24. Singh, L. K., Khanna, M., Thawkar, S. & Singh, R. Nature-inspired computing and machine learning based classification approach for glaucoma in retinal fundus images. *Multimed. Tools Appl.* **82**, 42851–42899 (2023).
25. Singh, L. K., Pooja, Garg, H. & Khanna, M. An IOT based predictive modeling for glaucoma detection in optical coherence tomography images using hybrid genetic algorithm. *Multimed. Tools Appl.* **81**, 37203–37242 (2022).
26. Singh, L. K. *et al.* Detection of glaucoma in retinal images based on multiobjective approach. *Int. J. Appl. Evolut. Comput. (IJAEC)* **11**, 15–27 (2020).
27. Lu, J. *et al.* Automated strabismus detection for telemedicine applications. Preprint at arXiv:1809.02940 (2018).
28. Kamal, M. M. *et al.* A comprehensive review on the diabetic retinopathy, glaucoma and strabismus detection techniques based on machine learning and deep learning. *Eur. J. Med. Health Sci.* **4**, 24–40 (2022).
29. Kaleem, S., Sohail, A., Tariq, M. U. & Asim, M. An improved big data analytics architecture using federated learning for IOT-enabled urban intelligent transportation systems. *Sustainability* **15**, 15333 (2023).
30. Suriyal, S., Druzgalski, C. & Gautam, K. Quantitative assessment of strabismus and selected vision related anomalies. In *World Congress on Medical Physics and Biomedical Engineering 2018: June 3–8, 2018, Prague, Czech Republic*. Vol. 1. 103–108 (Springer, 2019).
31. Jabbar, A. *et al.* Deep transfer learning-based automated diabetic retinopathy detection using retinal fundus images in remote areas. *Int. J. Comput. Intell. Syst.* **17**, 1–20 (2024).
32. Zhang, G. *et al.* Multi-feature fusion-based strabismus detection for children. *IET Image Process.* **17**, 1590–1602 (2023).
33. Chen, W. *et al.* Early detection of visual impairment in young children using a smartphone-based deep learning system. *Nat. Med.* **29**, 493–503 (2023).
34. Kang, Y. C. *et al.* Automated mathematical algorithm for quantitative measurement of strabismus based on photographs of nine cardinal gaze positions. *BioMed. Res. Int.* **2022** (2022).
35. Mengash, H. A. & Hosni Mahmoud, H. A. Methodology for detecting strabismus through video analysis and intelligent mining techniques. *Comput. Mater. Contin.* **67** (2021).

36. Siddique, A. A. *et al.* Covid-19 classification from x-ray images: An approach to implement federated learning on decentralized dataset. *Comput. Mater. Contin.* **75**, 3883–3901 (2023).
37. Zolkifli, N. S. & Nazari, A. Tracing of strabismus detection using Hough transform. In *2020 IEEE Student Conference on Research and Development (SCOReD)*, 313–318 (IEEE, 2020).
38. Angelov, P. & Gu, X. Mice: Multi-layer multi-model images classifier ensemble. In *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*, 1–8 (IEEE, 2017).
39. Mahmood, T., Rehman, A., Saba, T., Nadeem, L. & Bahaj, S. A. O. Recent advancements and future prospects in active deep learning for medical image segmentation and classification. *IEEE Access* (2023).
40. Peng, R. & Varshney, P. K. Noise-refined image enhancement using multi-objective optimisation. *IET Image Process.* **7**, 191–200 (2013).
41. Mahmood, T., Saba, T., Rehman, A. & Alamri, F. S. Harnessing the power of radiomics and deep learning for improved breast cancer diagnosis with multiparametric breast mammography. *Expert Syst. Appl.* 123747 (2024).
42. Jabbar, A. *et al.* Brain tumor detection and multi-grade segmentation through hybrid caps-vggnet model. *IEEE Access* (2023).
43. Mohammadi, K., Islam, A. & Belhaouari, S. B. Zooming into clarity: Image denoising through innovative autoencoder architectures. *IEEE Access* (2024).
44. Jabbar, A. *et al.* A lesion-based diabetic retinopathy detection through hybrid deep learning model. *IEEE Access* (2024).
45. Tan, T. Y. *et al.* Evolving ensemble models for image segmentation using enhanced particle swarm optimization. *IEEE Access* **7**, 34004–34019 (2019).
46. Madjarov, G., Gjorgjevikj, D. & Chorbev, I. A multi-class svm classifier utilizing binary decision tree. *Informatica* (2009).
47. Glasgow, M. R., Yuan, H. & Ma, T. Sharp bounds for federated averaging (local sgd) and continuous perspective. In *International Conference on Artificial Intelligence and Statistics*. 9050–9090 (PMLR, 2022).
48. Chatfield, K., Simonyan, K., Vedaldi, A. & Zisserman, A. Return of the devil in the details: Delving deep into convolutional nets. Preprint at arXiv:1405.3531 (2014).
49. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778 (2016).
50. Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations* (2020).
51. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*. 6105–6114 (2019).
52. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition* 4700–4708 (2017).
53. Howard, A. *et al.* Searching for mobilenetv3. In *Proc. of the IEEE/CVF International Conference on Computer Vision*. 1314–1324 (2019).
54. Nair, V. & Hinton, G. E. Rectified linear units improve restricted Boltzmann machines. In *Proc. of the 27th International Conference on Machine Learning (ICML-10)*. 807–814 (2010).
55. Zhang, X. *et al.* Attention to region: Region-based integration-and-recalibration networks for nuclear cataract classification using as-oct images. *Med. Image Anal.* **80**, 102499 (2022).
56. Zhang, X. *et al.* Regional context-based recalibration network for cataract recognition in as-oct. *Pattern Recognit.* **147**, 110069 (2024).
57. Miao, Y., Jeon, J. Y., Park, G., Park, S. W. & Heo, H. Virtual reality-based measurement of ocular deviation in strabismus. *Comput. Methods Prog. Biomed.* **185**, 105132 (2020).
58. Mao, K. *et al.* An artificial intelligence platform for the diagnosis and surgical planning of strabismus using corneal light-reflection photos. *Ann. Transl. Med.* **9** (2021).
59. Zhang, X. *et al.* Pyramid pixel context adaption network for medical image classification with supervised contrastive learning. *IEEE Transactions on Neural Networks and Learning Systems* (2024).
60. Li, J. *et al.* Evaluation of streamed hardware-to-software telemedicine strabismus consultations utilizing video glasses. *Clin. Ophthalmol. (Auckland, NZ)* **16**, 3927 (2022).
61. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556 (2014).

Acknowledgements
This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R513), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Authors are also thankful to Prince Sultan University, Riyadh Saudi Arabia for support.

Author contributions
Conceptualization, Ayesha Jabbar, Kashif Jabbar, Tariq Mahmood, Yasin and Tahani Jaser Alahmadi; Data Curation, Ayesha Jabbar, Haitham Nobanee, and Amjad Rehman; Formal analysis, Tariq, Amjad, Kashif and Haitham Nobanee; Funding acquisition, Tahani Jaser Alahmadi; Methodology, Ayesha Jabbar, Kashif Jabbar, Tariq Mahmood, Tahani Jaser Alahmadi and Amjad Rehman; Software, Ayesha Jabbar, Tariq Mahmood, Amjad Rehman, Yasin, and Haitham Nobanee; Supervision, Tariq Mahmood.

Funding
This research is funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R513), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Declarations

Competing interests
The authors declare no competing interests.

Institutional review board statement
All methods were carried out in accordance with relevant guidelines and regulations.

Additional information

Correspondence and requests for materials should be addressed to T.J.A.